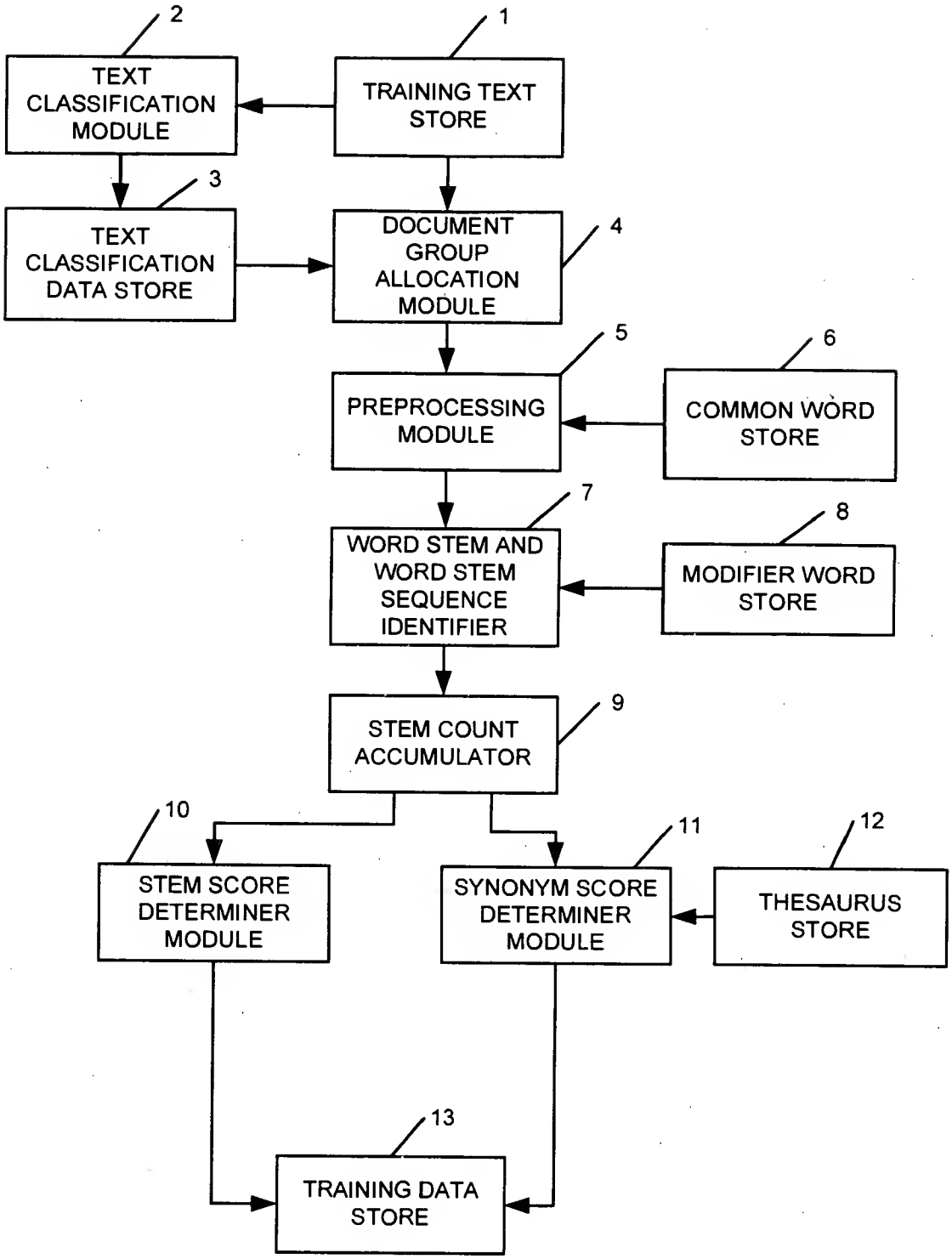
2
1
TEXT CLASSIFICATION MODULE
TRAINING TEXT STORE
3
TEXT CLASSIFICATION DATA STORE
DOCUMENT GROUP ALLOCATION MODULE
4
5
6
PREPROCESSING MODULE
COMMON WORD STORE
7
8
WORD STEM AND WORD STEM SEQUENCE IDENTIFIER
MODIFIER WORD STORE
STEM COUNT ACCUMULATOR
9
10
11
12
STEM SCORE DETERMINER MODULE
SYNONYM SCORE DETERMINER MODULE
THESAURUS STORE
13
TRAINING DATA STORE

Fig 1

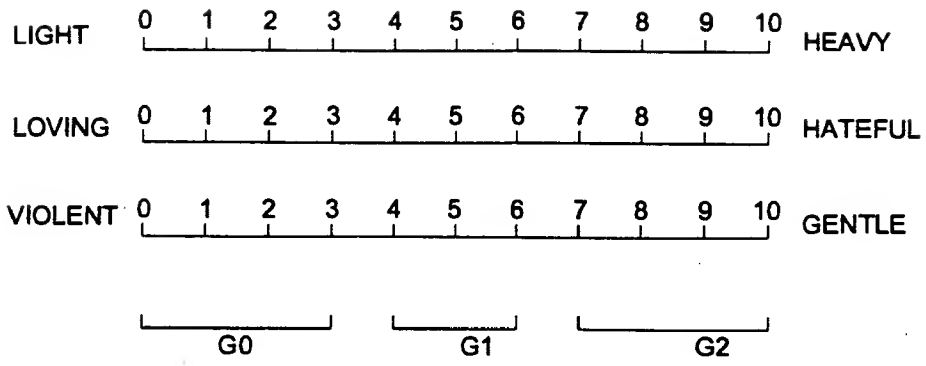
LIGHT
0 1 2 3 4 5 6 7 8 9 10
HEAVY
LOVING
0 1 2 3 4 5 6 7 8 9 10
HATEFUL
VIOLENT
0 1 2 3 4 5 6 7 8 9 10
GENTLE
G0
G1
G2

Fig 2

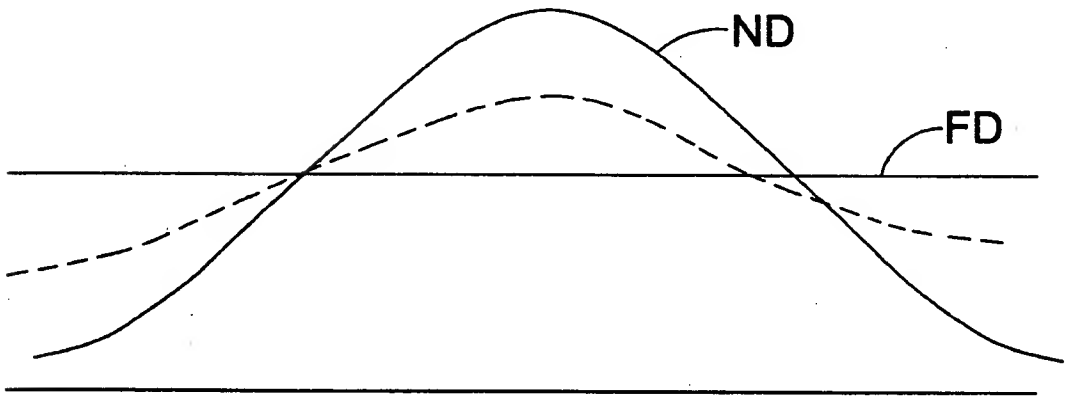
ND
FD

Fig 3

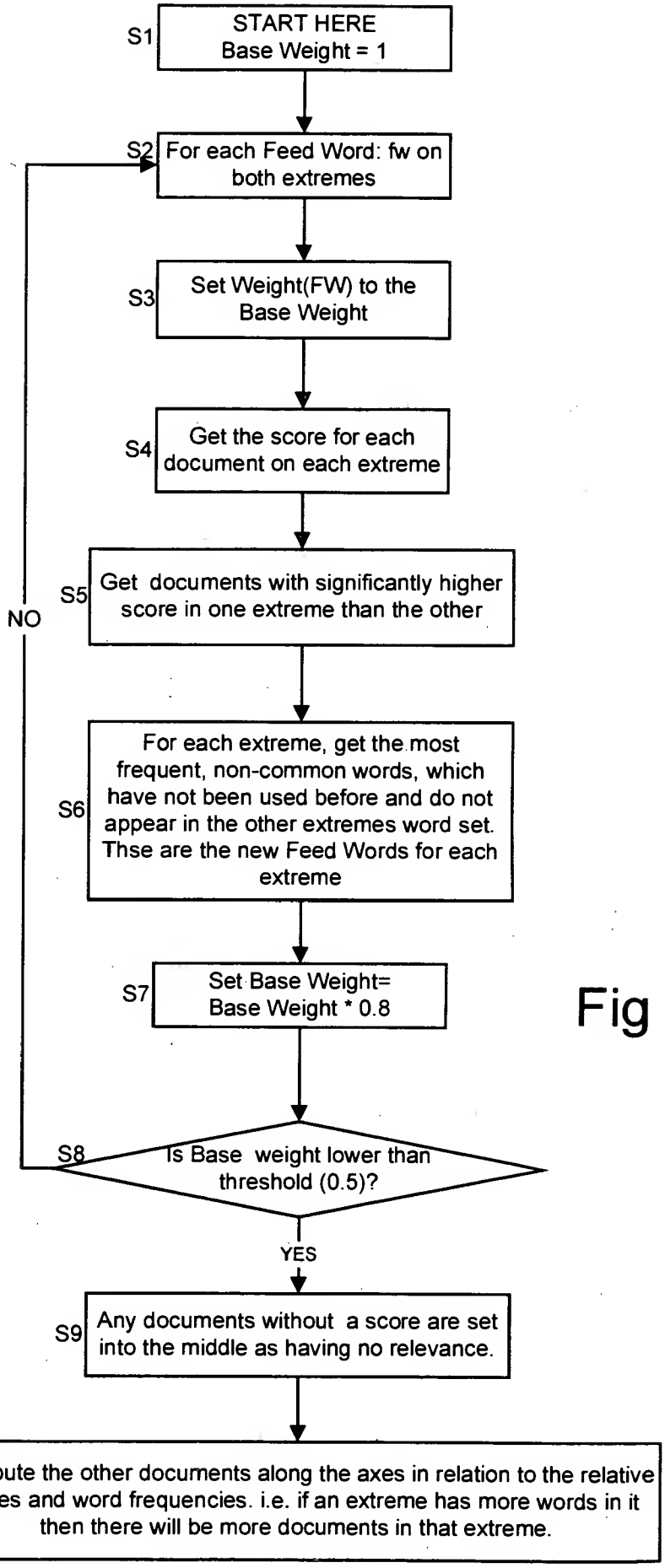
S1
START HERE
Base Weight = 1
S2
For each Feed Word: fw on both extremes
S3
Set Weight(FW) to the Base Weight
S4
Get the score for each document on each extreme
S5
Get documents with significantly higher score in one extreme than the other
NO
S6
For each extreme, get the most frequent, non-common words, which have not been used before and do not appear in the other extremes word set. Thse are the new Feed Words for each extreme
S7
Set Base Weight= Base Weight * 0.8
S8
Is Base weight lower than threshold (0.5)?
YES
S9
Any documents without a score are set into the middle as having no relevance.
S10
Distribute the other documents along the axes in relation to the relative scores and word frequencies. i.e. if an extreme has more words in it then there will be more documents in that extreme.

Fig 4a

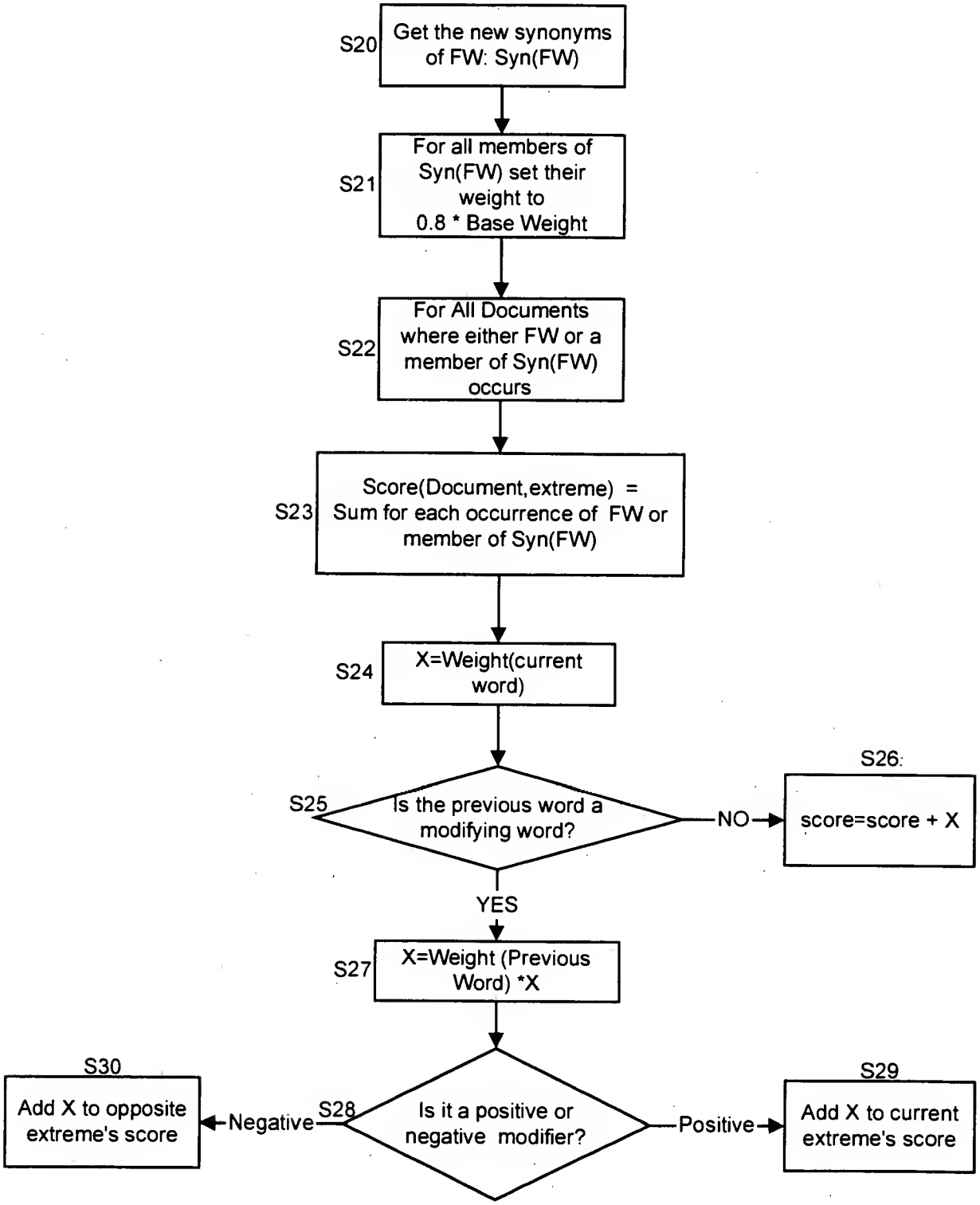
S20
Get the new synonyms of FW: Syn(FW)
S21
For all members of Syn(FW) set their weight to 0.8 * Base Weight
S22
For All Documents where either FW or a member of Syn(FW) occurs
S23
Score(Document,extreme) = Sum for each occurrence of FW or member of Syn(FW)
S24
X=Weight(current word)
S25
Is the previous word a modifying word?
NO
S26:
score=score + X
YES
S27
X=Weight (Previous Word) *X
S28
Is it a positive or negative modifier?
Negative
S30
Add X to opposite extreme's score
Positive
S29
Add X to current extreme's score

Fig 4b

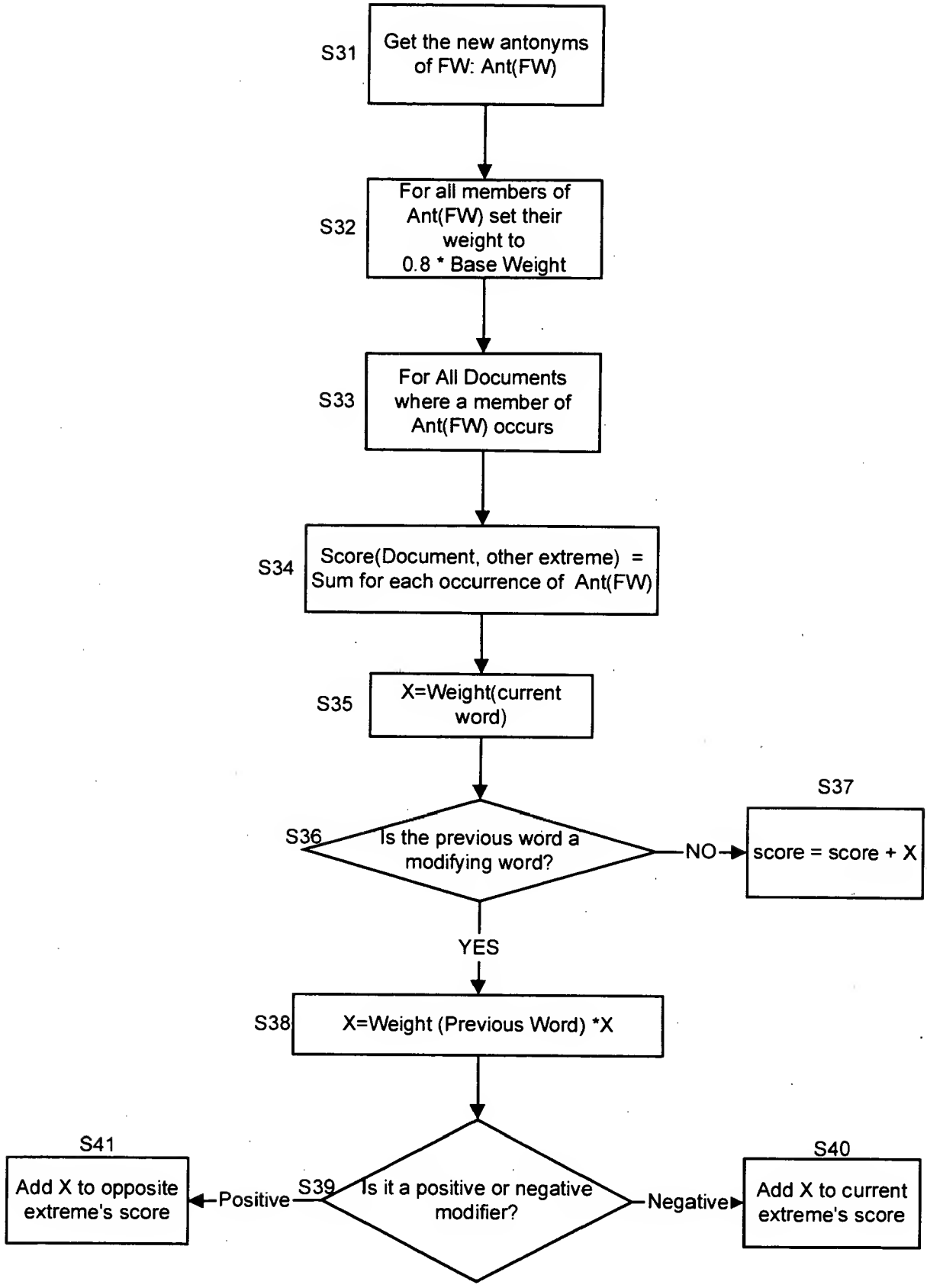
S31
Get the new antonyms of FW: Ant(FW)
S32
For all members of Ant(FW) set their weight to 0.8 * Base Weight
S33
For All Documents where a member of Ant(FW) occurs
S34
Score(Document, other extreme) = Sum for each occurrence of Ant(FW)
S35
X=Weight(current word)
S36
Is the previous word a modifying word?
NO
S37
score = score + X
YES
S38
X=Weight (Previous Word) *X
S39
Is it a positive or negative modifier?
Positive
S41
Add X to opposite extreme's score
Negative
S40
Add X to current extreme's score

Fig 4c

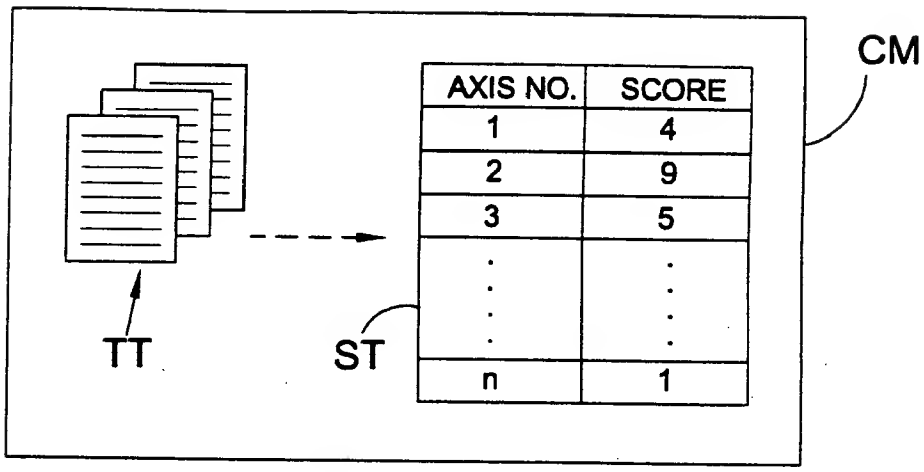

Fig 5

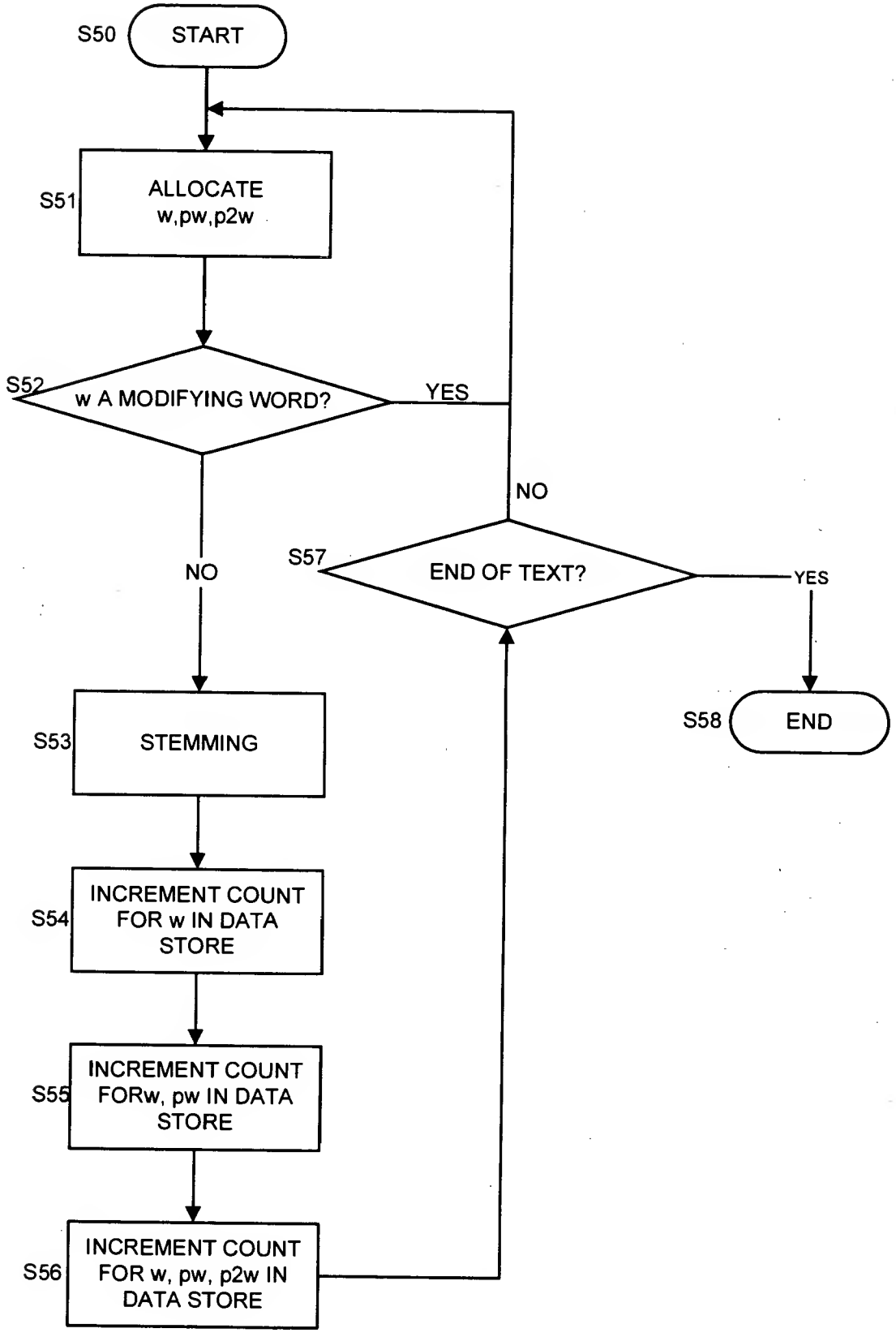
S50
START
S51
ALLOCATE
w,pw,p2w
S52
w A MODIFYING WORD?
YES
NO
S57
END OF TEXT?
YES
NO
S58
END
S53
STEMMING
S54
INCREMENT COUNT
FOR w IN DATA
STORE
S55
INCREMENT COUNT
FORw, pw IN DATA
STORE
S56
INCREMENT COUNT
FOR w, pw, p2w IN
DATA STORE

Fig 6

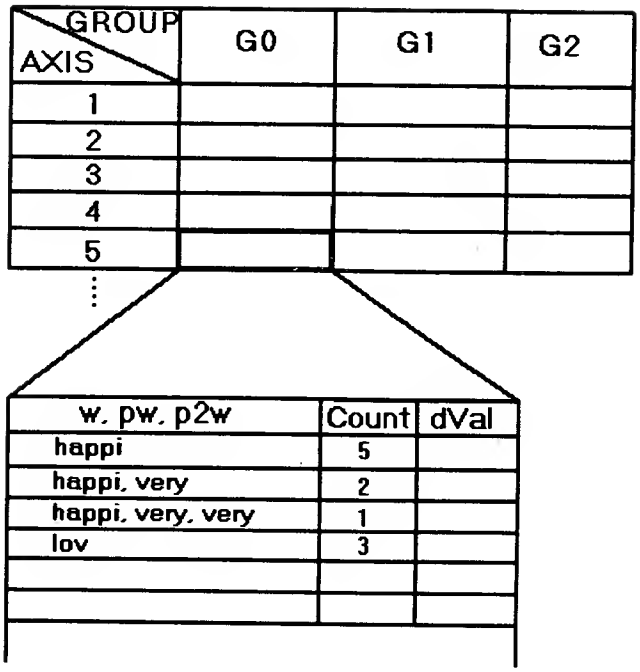
GROUP
AXIS
G0
G1
G2
1
2
3
4
5
w, pw, p2w
Count
dVal
happi
5
happi, very
2
happi, very, very
1
lov
3

Fig 7

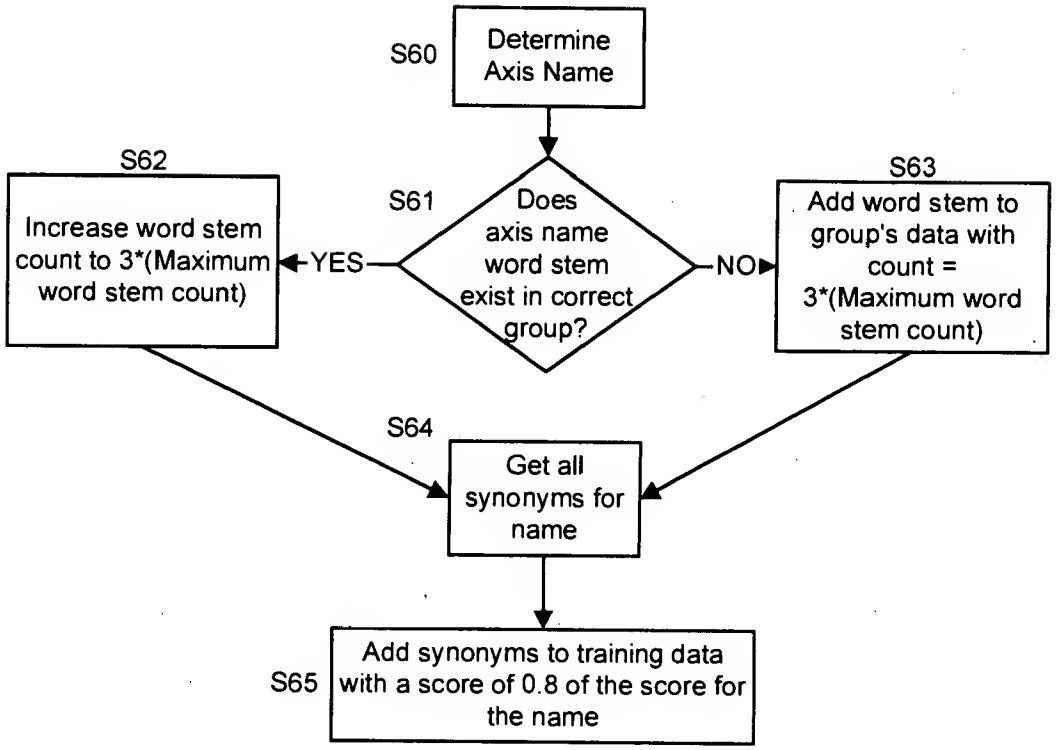
S60
Determine Axis Name
S61
Does axis name word stem exist in correct group?
YES
NO
S62
Increase word stem count to 3*(Maximum word stem count)
S63
Add word stem to group's data with count = 3*(Maximum word stem count)
S64
Get all synonyms for name
S65
Add synonyms to training data with a score of 0.8 of the score for the name

Fig 8

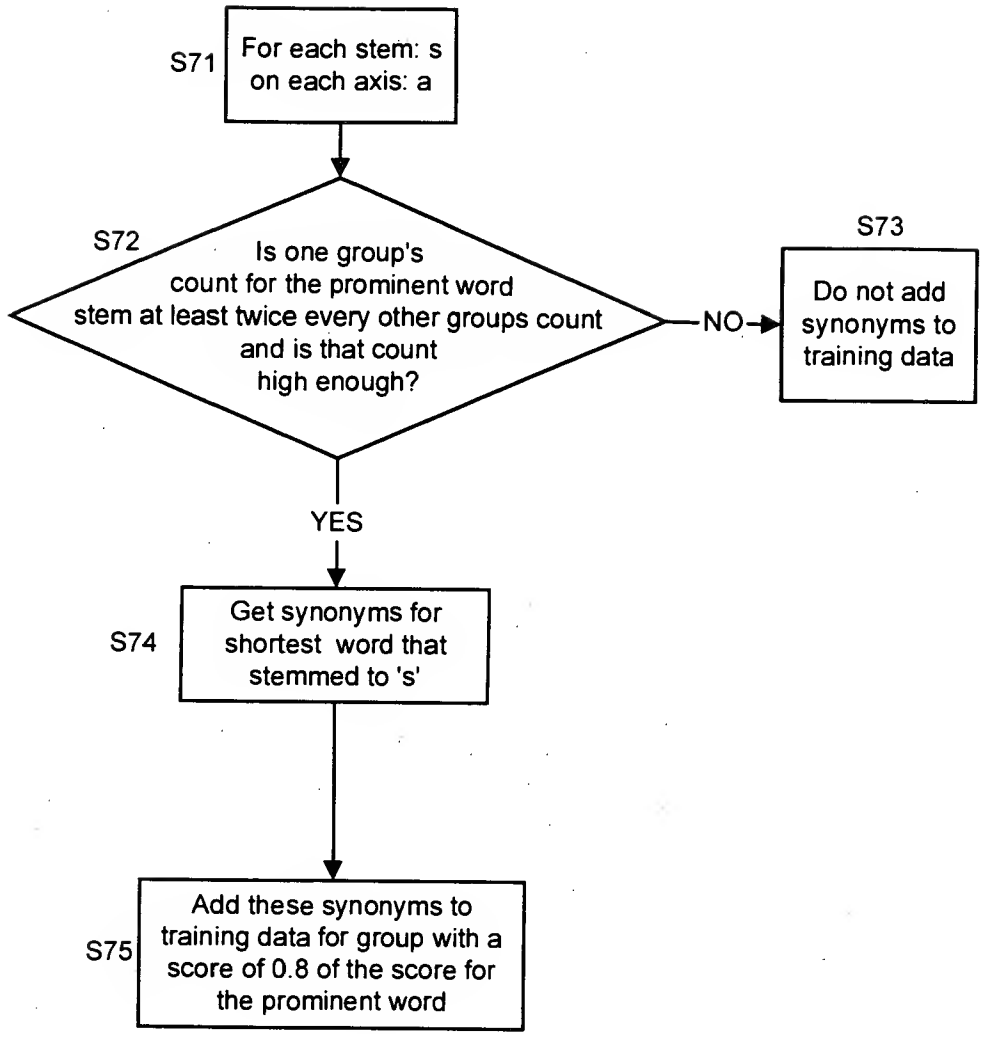
S71
For each stem: s on each axis: a
S72
Is one group's count for the prominent word stem at least twice every other groups count and is that count high enough?
NO
S73
Do not add synonyms to training data
YES
S74
Get synonyms for shortest word that stemmed to 's'
S75
Add these synonyms to training data for group with a score of 0.8 of the score for the prominent word

Fig 9

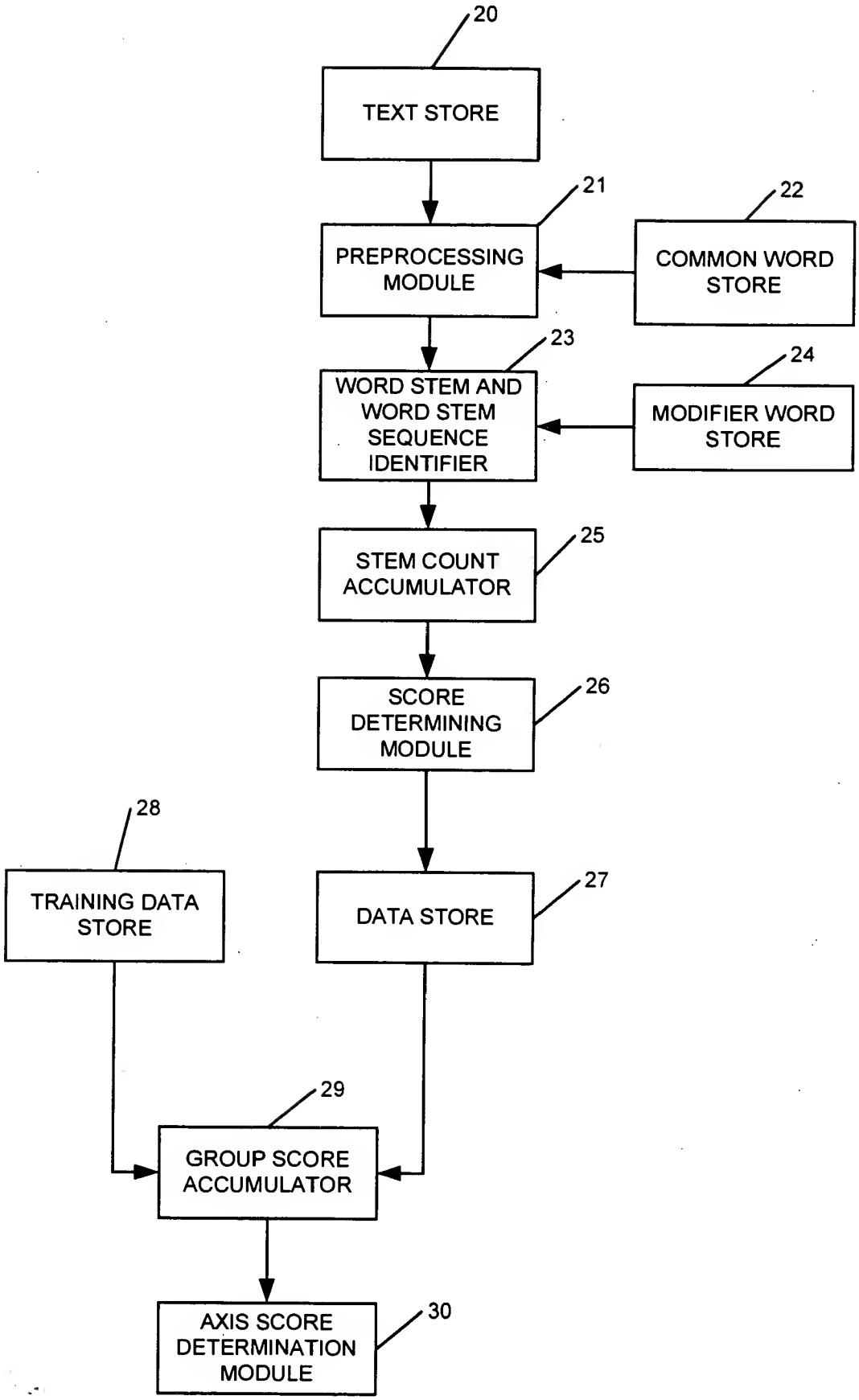
20
TEXT STORE
21
PREPROCESSING MODULE
22
COMMON WORD STORE
23
WORD STEM AND WORD STEM SEQUENCE IDENTIFIER
24
MODIFIER WORD STORE
25
STEM COUNT ACCUMULATOR
26
SCORE DETERMINING MODULE
28
TRAINING DATA STORE
27
DATA STORE
29
GROUP SCORE ACCUMULATOR
30
AXIS SCORE DETERMINATION MODULE

Fig 10

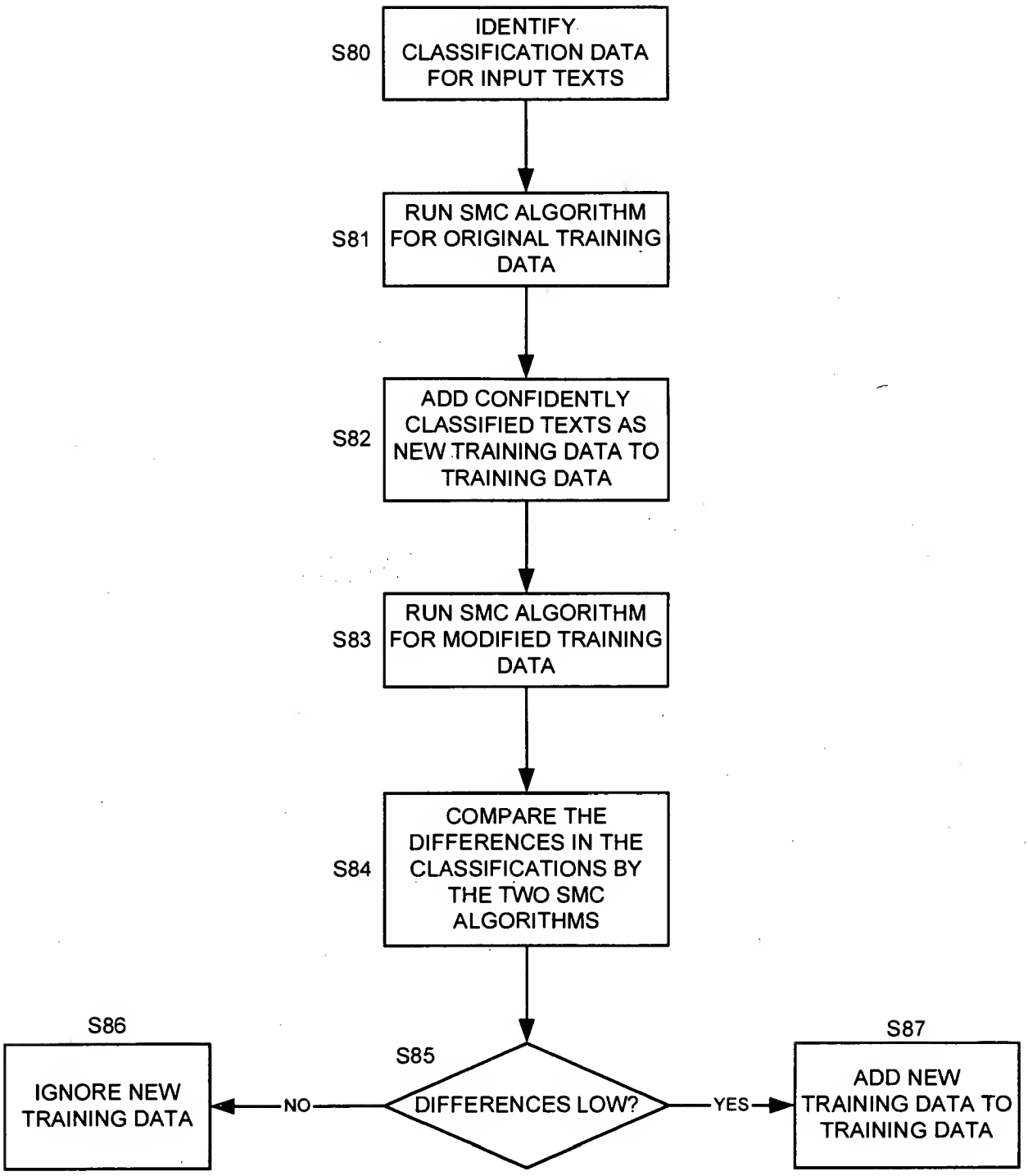
S80
IDENTIFY CLASSIFICATION DATA FOR INPUT TEXTS
S81
RUN SMC ALGORITHM FOR ORIGINAL TRAINING DATA
S82
ADD CONFIDENTLY CLASSIFIED TEXTS AS NEW TRAINING DATA TO TRAINING DATA
S83
RUN SMC ALGORITHM FOR MODIFIED TRAINING DATA
S84
COMPARE THE DIFFERENCES IN THE CLASSIFICATIONS BY THE TWO SMC ALGORITHMS
S85
DIFFERENCES LOW?
NO
YES
S86
IGNORE NEW TRAINING DATA
S87
ADD NEW TRAINING DATA TO TRAINING DATA

Fig 11

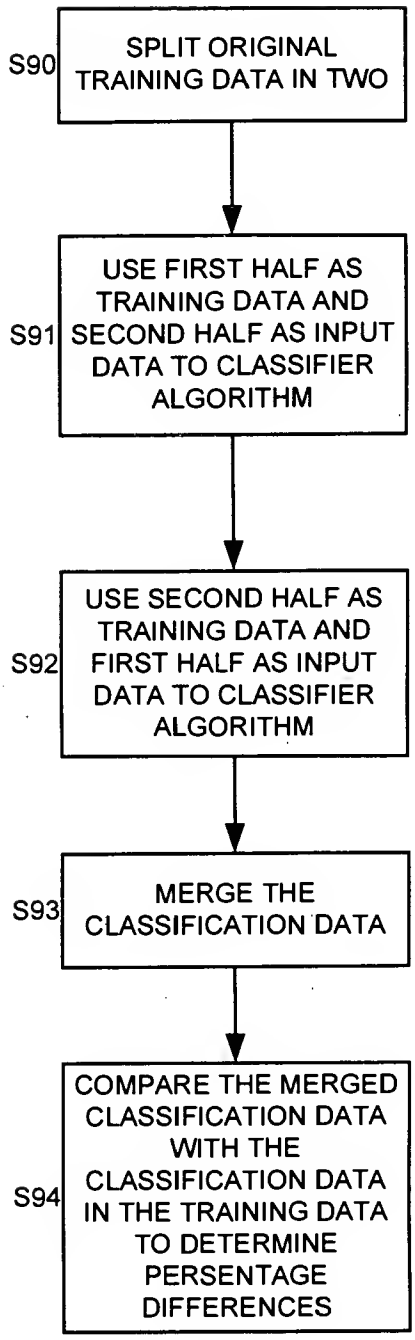
S90
SPLIT ORIGINAL TRAINING DATA IN TWO
S91
USE FIRST HALF AS TRAINING DATA AND SECOND HALF AS INPUT DATA TO CLASSIFIER ALGORITHM
S92
USE SECOND HALF AS TRAINING DATA AND FIRST HALF AS INPUT DATA TO CLASSIFIER ALGORITHM
S93
MERGE THE CLASSIFICATION DATA
S94
COMPARE THE MERGED CLASSIFICATION DATA WITH THE CLASSIFICATION DATA IN THE TRAINING DATA TO DETERMINE PERSENTAGE DIFFERENCES

Fig 12

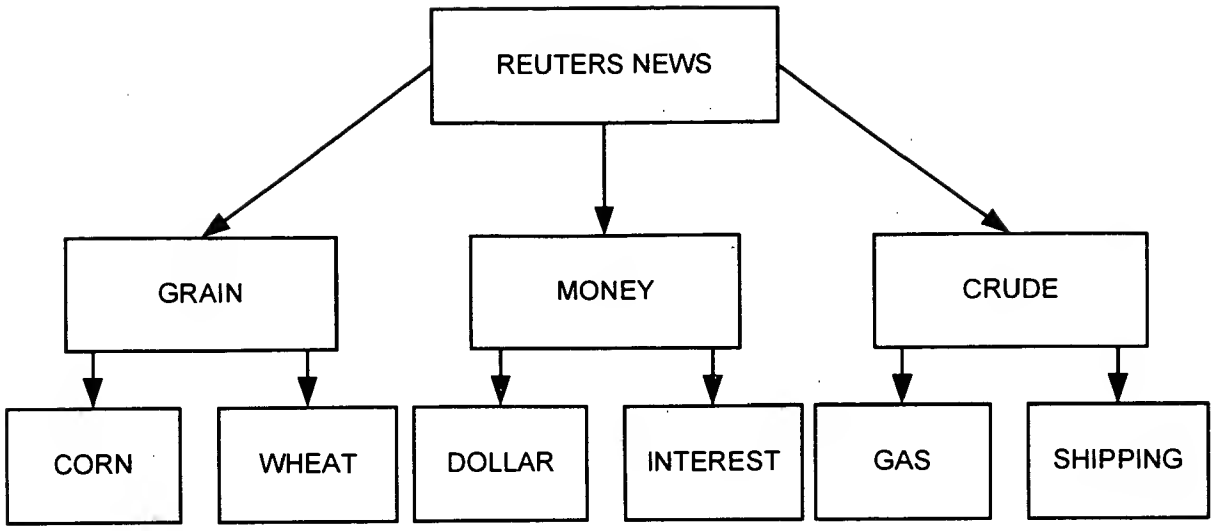
REUTERS NEWS
GRAIN
MONEY
CRUDE
CORN
WHEAT
DOLLAR
INTEREST
GAS
SHIPPING

Fig 13

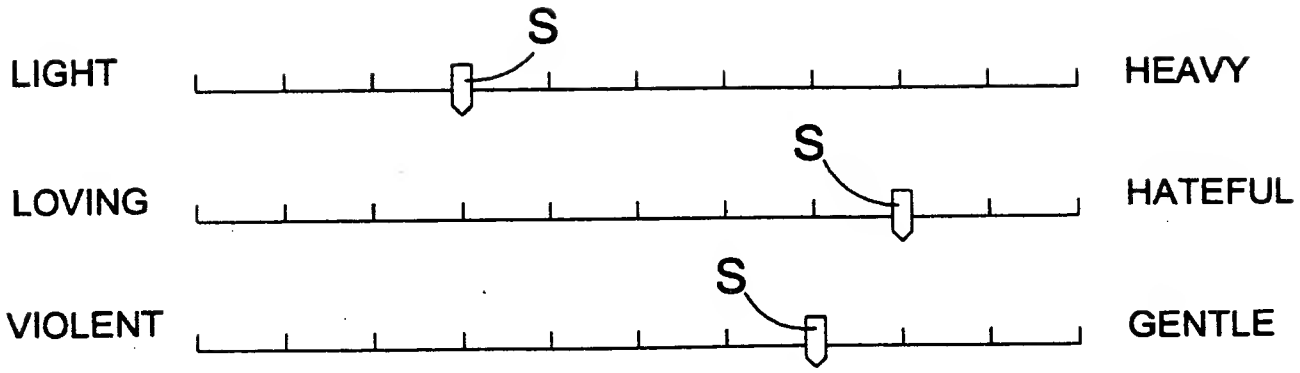
LIGHT
S
HEAVY
LOVING
S
HATEFUL
VIOLENT
S
GENTLE

Fig 14

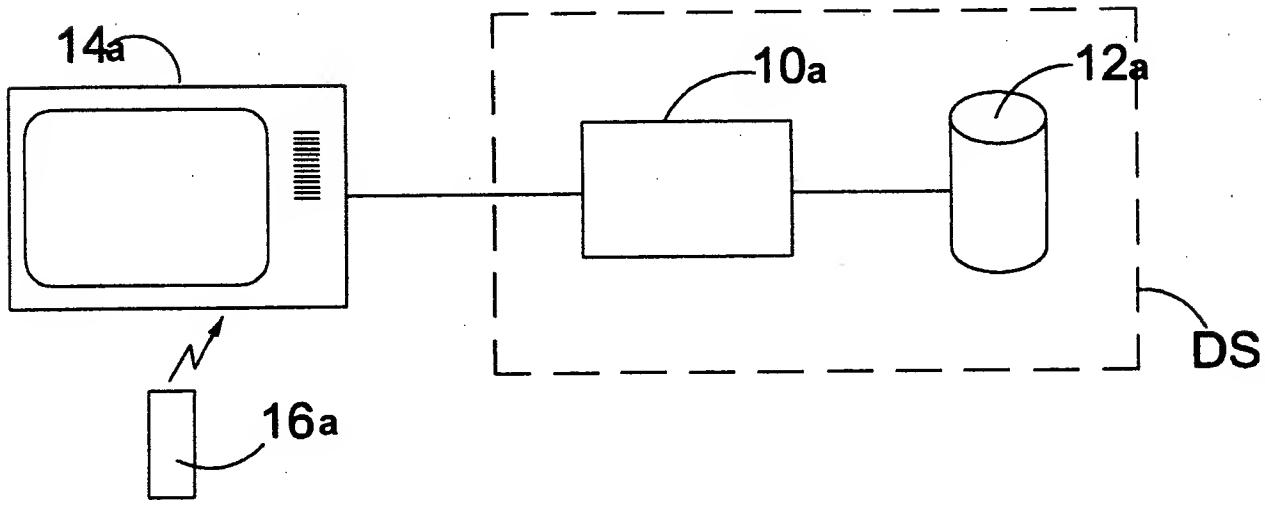
14a
10a
12a
DS
16a

Fig 15